Accepted for publication in: ***Journal of Cognitive Neuroscience***

Parallel Processing of Semantics and Phonology in Spoken Production: Evidence from Blocked Cyclic Picture Naming and EEG

Chen Feng[1 2], Markus F. Damian[3], Qingqing Qu[1 2]

[1]Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China
[2]Department of Psychology, University of Chinese Academy of Sciences, Beijing, China
[3]School of Psychological Science, University of Bristol, United Kingdom

Word count: 7,426 (main text, excl. references)

Address for correspondence:

Qingqing Qu
Key Laboratory of Behavioral Science
Institute of Psychology
Chinese Academy of Sciences, Beijing, China
16 Lincui Road, Chaoyang District, Beijing, China
100101
China
Tel: +86-10-64888629
Fax: +86-10-64872010

Email: quqq@psych.ac.cn

Abstract

Spoken language production involves lexical-semantic access and phonological encoding. A theoretically important question concerns the relative time course of these two cognitive processes. The predominant view has been that semantic and phonological codes are accessed in successive stages. However, recent evidence seems difficult to reconcile with a sequential view but rather suggests that both types of codes are accessed in parallel. Here, we used event-related potentials (ERPs) combined with the 'blocked cyclic naming paradigm' in which items overlapped either semantically or phonologically. Behaviourally, both semantic and phonological overlap caused interference relative to unrelated baseline conditions. Crucially, ERP data demonstrated that the semantic and phonological effects emerged at a similar latency (~180 ms after picture onset) and within a similar time window (180-380 ms). These findings suggest that access to phonological information takes place at a relatively early stage during spoken planning, largely in parallel with semantic processing.

Keywords: ERPs; Phonology; Semantics; Spoken production; Time Course

Introduction

An important issue in psychology in general, and psycholinguistics in particular, is how multiple cognitive processes take place across time and with regard to one another. Spoken language production involves lexical-semantic and phonological encoding. At the former level, a cohort of semantically related lexical nodes are activated and a lexical target node is selected among co-activated competitors, while at the latter level, phonological forms of words are accessed which enables the following articulation. How cognitive action takes place across semantic and phonological encoding has been one of the hotly debated topics in the domain of language production.

Traditionally, the predominant view held that semantic and phonological codes are processed in two sequential steps, with access to semantic codes preceding phonological processing, and semantic processing is completed prior to phonological processing. This assumption is embedded in “discrete” or ”serial” models of production, such as those by Garrett (1975) and Levelt, Roelofs and Meyer (1999). However, various findings from the literature on speech errors, as well as from experimental tasks which elicit spoken responses in the laboratory have cast doubt on the assumption of strict seriality, and hence a number of formal models of spoken production have introduced some degree of “interactivity” between processing levels, with the most prominent interactive model proposed by Dell (1986; Dell, Schwartz, Martin, Saffran, & Gagnon, 1997). Rapp and Goldrick (2000) provide a detailed theoretical unfolding of the notions of seriality, interactivity, cascadedness, and feedback. Even with interactive principles, most models still adhere to a broad notion of sequentiality, such that in the preparation of a spoken utterance, semantic codes are accessed first, whereas phonological properties are retrieved later. By contrast, a number of recent studies have reported findings on the basis of which the authors have challenged this prominent sequential view and have instead postulated a parallel account according to which speakers rapidly access semantic and phonological information largely in parallel (e.g., Miozzo, Pulvermüller, & Hauk, 2015; Strijkers, Costa, & Pulvermüller, 2017). An exploration of the relative time course of the two cognitive processes is therefore theoretically important as it would provide a critical test of sequential versus parallel processing. In this study, we investigated the issue by tracking the temporal dynamics of semantic and phonological processes in spoken production via high time-resolution electrophysiological measurement, i.e., event-related potentials (ERPs).

Early electrophysiological investigations on the time course of semantics and phonology employed covert (rather than overt) naming tasks, and largely based on this evidence, Indefrey and Levelt (2004) provided estimates of temporal windows corresponding to the processes underlying word production. Spoken production was described as a serial succession of processes elicited by a single-word production episode such as in object naming, with semantic processing taking place from 200-275 ms, and phonological encoding from 275-450 ms after picture onset. However, there is increasing concern that from a methodological perspective, these early studies were not ideally suited to detect the time course of the underlying processes. For example, studies relying on motor response preparation or inhibition indexed by ‘lateralized readiness potentials’ (LRPs) and N200 in semantic or phonological classification tasks suggested that the availability of semantic information precedes that of phonological information (Schmitt, Münte, & Kutas, 2000; Van Turennout, Hagoort, & Brown, 1997; but see Abdel Rahman & Sommer, 2003). As these early ERP studies did not involve actual overt speech production, but rather relied on complex meta-linguistic decisions linked to button-press responses (e.g., semantic-phonological categorization task in van Turennout et al. 1997), they could reflect response decision rather than online timing of cognitive processes underlying a naming response (see Strijkers & Costa, 2011). Moreover, it has been argued that LRPs or N200 are only informative as to the termination of semantic and phonological processes but not their initiation (Abdel Rahman & Sommer, 2003; Camen, Morand, & Laganaro, 2010; Laganaro & Perret, 2011). Finally, the complexity associated with these experimental tasks allows for several alternative interpretations (see Strijkers & Costa, 2016 for a review).

More recent studies have successfully combined overt spoken production with EEG measurements. A particularly rich source of empirical evidence for overt spoken production has been derived from the picture-word interference (PWI) task. Relative to a condition in which target picture and distractor word are unrelated, a semantic relationship slows down naming, whereas a phonological relationship speeds up responses (Glaser & Düngelhoff, 1984; Schriefers, Meyer, & Levelt, 1990; Starreveld & La Heij, 1995). This semantic interference has arguably been assumed to arise at the stage of lexical-semantic retrieval and phonological facilitation arises at the stage of phonological encoding. The current results clearly demonstrate a statistical interaction between semantic and phonological relatedness in PWI tasks, which were taken as evidence for the non-serial view (Damian & Martin, 1999; Starreveld & La Heij, 1995, 1996; Taylor & Burke, 2002). Dell'Acqua et al. (2010) used ERPs combined with the PWI task to track the time course of semantic and phonological encoding and found sequential time windows for the semantic and phonological effects (see Zhu et al. 2015 for the similar findings in Mandarin spoken production). However, concerns about the picture-word interference task for investigating the time course of production have been voiced (Strijkers & Costa, 2011). One possible criticism is that the exact locus of semantic effects from the picture-word interference task remains controversial. For example, semantic effects have been attributed to semantic-lexical level (e.g., Glaser & Glaser, 1982; Levelt et al., 1999; Roelofs, 1992; Schriefers et al., 1990; Starreveld & La Heij, 1996), or alternatively, a postlexical articulatory level by the "response exclusion" hypothesis (e.g., Mahon, Costa, Peterson, Vargas, & Caramazza, 2007). Moreover, the superimposition of a visually complex distractor could itself delay the cognitive processes associated with object name production (see Qu &Damian 2020 for a review).

A number of recent studies have reported findings which suggest a parallel time course of semantics and phonology, a pattern which is difficult to reconcile with a sequential model which would predict more sequential time signatures (e.g., Miozzo et al., 2015; Strijkers, Costa, & Thierry, 2010; Strijkers et al., 2017; for reviews see Munding, Dubarry, & Alario, 2016; Strijkers & Costa, 2016). First, it has been demonstrated that phonological manipulations modulated ERPs starting around 200 ms after picture presentation. For example, Qu, Damian, & Kazanina (2012) asked Chinese speakers to name coloured objects with adjective-noun phrases, and found that phoneme overlap between adjective and noun modulated ERPs starting from 200 ms after picture onset (see also Qu, Feng, Hou, & Damian, 2020; Yu, Mo & Mo, 2014, for a similar finding with various tasks). Such an early onset of phonological effects in spoken production would be incompatible with the time estimates provided by Indefrey & Levelt (2004; Indefrey, 2011), and in fact would imply that semantic and phonological processing begins roughly simultaneously. Using a multiple linear regression approach in an MEG study of picture naming, Miozzo et al. (2015) manipulated variables which were assumed to primarily impact semantic (semantic features and action features) and phonological processing (word length, phonological neighborhood size). They found an early and simultaneous latency of semantic and phonological effects at around 150 ms upon object presentation. In another study, Strijkers et al. (2017) went beyond "adjacent" processing layers (lexical-semantic vs. phonological encoding) and investigated the interactivity of a more extreme contrast targeting the lexical-semantic and phonetic-articulatory processes in object naming. Lexical-semantic processing was assessed by manipulating word frequency, while phonetic-articulatory properties were manipulated by the place of articulation of word-initial phonemes (lip vs. tongue, e.g., Monkey vs. Donkey). Brain responses which were modulated by word frequency of object names emerged in 160-240 and 260-340 ms, while lip-tongue contrasts were associated with 160-240 ms. Overall, recent results such as these highlight an early access to phonological codes in spoken production, and hence constitute a challenge for most traditional models which have generally assumed that phonology is accessed later than semantics.

The present study

In the current study, rather than tracking relevant psycholinguistic variables such as semantic features or word frequency (as in Miozzo et al., 2015), we attempted to assess the relative time course of semantic and phonological access via the "blocked cyclic naming" paradigm, in which participants repeatedly name small sets of pictures in short experimental blocks, and the semantic or phonological context for a given block is manipulated. In its semantic form, this task is commonly used to explore lexical-semantic retrieval (Abdel Rahman & Melinger, 2007; Aristei, Melinger, & Abdel Rahman, 2011; Damian, Vigliocco, & Levelt, 2001; Kroll & Stewart 1994). Pictures are arranged such that within a block, all pictures stem from a single semantic category ("homogeneous"), or from various categories ("heterogeneous"). Across the study, each picture is named both in homogeneous and heterogeneous contexts, and hence acts as its own control; the only aspect which varies is the homogeneous/heterogeneous semantic context. The typical finding is that objects are named slower in homogeneous than in heterogeneous blocks (Abdel Rahman & Melinger, 2009; Damian et al., 2001; Kroll & Stewart, 1994). This semantic blocking effect has been argued to originate at the lexical-semantic processing stage, although the mechanism underlying the semantic blocking effect is debated (e.g., Damian & Als, 2005; Damian et al., 2001; Howard, Nickels, Coltheart, & Cole-Virtue, 2006; Oppenheim, Dell, & Schwartz, 2010).

Rather than manipulating semantic context, one can also manipulate phonological context, such as whether or not responses within an experimental block share phonological properties. Here, the general finding is that word-initial overlap among items within a block facilitates responses (e.g., Chen & Chen, 2013; Damian, 2003; Meyer, 1990, 1991; Roelofs, 2006; O'Seaghdha, Chen, & Chen, 2010;) whereas non-initial overlap does not yield a benefit (e.g., Meyer, 1990, 1991). This phonological facilitation has conventionally been explained through partial planning of the response operating at the phonological processing level which is possible in homogeneous but not in heterogeneous blocks (Roelofs, 1997). However, this explanation has recently been called into question by O'Séaghdha and Frazer (2014) who highlighted attentional influences on the effect, which raise the possibility that the effect resides outside the language system proper. In a recent paper, Breining, Nozari, & Rapp (2016) reported highly intriguing results obtained with this task which pose a further challenge with regard to how the effect informs us with regard to language production. As in numerous previous articles, they blocked response words within a set by form overlap. However, contrary to being limited to the word-initial position in previous studies, in the homogeneous blocks responses were chosen such that overlap was distributed unpredictably across positions in words (e.g., cat-mat-cot-cap-map-mop). The critical finding was that this form of overlap slowed down naming latencies, compared to "heterogeneous" sets. This "similarity-based interference" effect is hence similar to the interference effect arising from semantic blocking summarised above, and the authors attributed both effects to a similar underlying principle of "incremental learning" (this notion will be unfolded in more detail in the Discussion).

In the current study, we manipulated semantic and phonological context within the same study and combined it with measurement of EEG. EPRs associated with the two types of context can provide important insights into the time course underlying word production. Our participants were native speakers of Chinese Mandarin. Semantic context was manipulated via blocking of semantic category membership. For the phonological manipulation, all picture names within a homogeneous block shared an atonal syllable but for half of the items the overlapping syllable was in the first position whereas for the other half it was in second position. To the best of our knowledge, our study constitutes the first attempt to track the time course of phonological encoding with ‘inconsistent’ overlap via EEG, and to compare it directly with the semantic context effect. Behaviourally, both semantic and phonological context should result in interference relative to the unrelated baseline conditions. In EEG, we expected the semantic effect to begin at a relatively early time; as shown in Table 1, in similar studies the semantic effect began around 200 ms post picture onset (the mean latency is 210 ms). The critical question was the onset of the phonological, relative to the semantic, effect. Models of spoken production which assume some degree of “sequentiality” between semantic and phonological stages would predict a later onset for phonological than for semantic effects. However, if both effects begin similarly early, that would constitute further difficulties for these accounts, and favour a notion according to which access to semantic and to phonological codes in spoken production can occur in parallel.

Methods

Participants

Following previous studies (e.g., Breining et al., 2016; Damian & Bowers, 2003), twenty-four native Mandarin Chinese speakers (17 females, mean age 22 years) participated and were compensated for their time. All participants were right-handed, with normal or corrected-to -normal version and no history of language disorders. Participants gave informed consent and the study was approved by the ethics committee of the Institute of Psychology, Chinese Academy of Sciences.

Materials and Design

For the semantic “type of overlap”, 16 objects were selected from four semantic categories and were arranged in a 4 × 4 matrix so that items in rows formed semantically homogeneous sets (i.e., items within a row stemmed from the same semantic category) whereas items in columns formed semantically heterogeneous sets (all items were from different categories). Items within a semantic homogeneous set were selected to minimise within-category visual similarity (e.g., 帆船, /fan1chuan2/, “steamboat”; 轿车, /jiao4che1/, “car”; 飞机, /fei1ji1/, “airplane”; 摩托, /mo2tuo2/, “motorbike”).

For the phonological “type of overlap”, another set of 16 objects were chosen such that in phonologically homogeneous sets, all four items shared an atonal syllable but for half items the overlapping syllable was in word-initial position whereas for the other half it was in word-final position (e.g., 石头, /<u>shi</u>2tou0/, “rock”; 试管, /<u>shi</u>4guan3/, “tube”; 钥匙, /yao4<u>shi</u>0/, “key”; 电视, /dai4<u>shi</u>4/, “television”). For the corresponding phonologically heterogeneous sets, phonological overlap was avoided (e.g., 颜料, /yan2liao4/, “paint”; 帽子, /mao4zi0/, “hat”; 石头, /shi2tou0/, “rock”; 鲸鱼, /jing1yu2/, “whale”).

Table 1. *Summary of studies concerning semantic context effect*

Study	Technique	Time Window (ms)	Language
Anders et al. (2019)	MEG	180-1000	French
Aristei et al. (2011)	EEG	200-550	German
Janssen et al. (2011)	EEG	220-450	Spanish
Janssen et al. (2015)	EEG	250-400 & 500-750	Spanish
Maess et al. (2002)	MEG	150-225	German
Python et al. (2018)	EEG	270-315	French
Wang et al. (2018)	EEG	200-550	Mandarin

For the semantic combinations, phonological overlap was minimised; for the phonological combinations, semantic overlap was minimised. In all sets, orthographic overlap between items was avoided. Across the semantic and phonology conditions, pictures were statistically matched on various lexical properties.[1] A full list of the materials is provided in Appendix A as supplementary data.

Type of overlap (semantics vs. phonology) was manipulated as a within-participant and between-item variable, and context (homogeneous vs. heterogeneous) was varied as within-participant and within-item variable. Half of the participants received the eight semantic blocks first and the remaining half received the eight phonology blocks first. Further, eight blocks (four homogeneous and four heterogeneous) were presented in an alternating sequence. The order of the four blocks was determined by a Latin Square design. Within each block, each item was presented for four cycles, resulting in 16 trials, in a pseudo-random order such that items were never repeated on adjacent trials. For each participant, the entire testing session included 256 trials (16 trials in each of 16 blocks).

Procedure

The experiment was run using the E-Prime 3.0 software (Psychology Software Tools, Pittsburgh, PA), with a microphone connected to the computer recording vocal responses. Participants were tested individually and were instructed to name objects as fast and accurately as possible. Prior to each block, participants were first asked to familiarise themselves with the four pictures for that block, with the corresponding names printed underneath each object. Each trial started with a fixation (500 ms) and then a blank screen (500 ms) was followed by an object (2,000 ms) in the center of the screen against a white background. The intertrial interval was 1,000 ms. Participants received a practice block comprising four filler objects, followed by the 16 experimental blocks. The entire experiment lasted approximately 90 minutes per participant.

EEG Recordings and Pre-processing

EEG signals were recorded with 64 electrodes secured in an elastic cap (Electro Cap International) using Neuroscan 4.3. The vertical electrooculogram (VEOG) was monitored with electrodes placed above and below the left eye. The horizontal EOG (HEOG) was recorded by a bipolar montage using two electrodes

[1] Stimuli in the semantic and phonological conditions were matched on the following variables: word frequency (p = .081), visual complexity (p = .961), naming agreement (p = .602), and stroke number (p = .322). Values were taken from Liu, Yao, Li, & Shu (2011).

placed on the right and left external cantus. The left mastoid electrode served as a reference. All electrode impedances were below 5 kΩ. Electrophysiological signals were amplified with a band-pass filter of 0.05 and 70 Hz (sampling rate 500 Hz).

The EEGLAB toolbox based on MATLAB was used for the following procedure of preprocessing EEG signals: offline filter with a high-pass cutoff of 0.1 Hz and a low-pass cutoff of 30 Hz, removal of ocular, muscle, motor artefacts and linear noise using independent component analyses (ICA; Jung et al., 2000) on the segmented data (−0.8 s to 1.5 s relative to the picture onset), manual rejection of epochs with extensive fluctuation and signals below/above ±70 μV, offline re-referencing against the average reference. The EEG was segmented into 600 ms epochs relative to picture onset that included a 100 ms pre-stimulus baseline and a 500 ms post-stimulus interval.

Response Latency Analysis

Naming latencies were analysed using a linear mixed-effects model (Baayen, Davidson, & Bates, 2008) with the package *lme4* (Bates & Maechler, 2009). We constructed a model including the main effects of type of overlap, context, and cycle, as well as their interactions (*model. matrix = ~type*context*cycle*). Further, to evaluate the context effects along with cycles within each type, another model was constructed by including the main effect of type, and simple effects of context, cycle and interactions between context and cycle in each type (*model. matrix = ~type/(context*cycle)*). In each model, a full random structure was implemented, with random intercepts and random slopes over participants and items. When the model failed to converge, a reduction procedure with the principle component analysis was conducted (Bates, Kliegl, Vasishth, & Baayen, 2015) with the rePCA function in the package *RePsychLing* (Baayen, Bates, Kliegl, & Vasishth, 2015) until the model could be supported by the data. We report *p* values derived from the *t* values using the normal approximation (Mirman, 2014).

ERP Analysis

Two types of analyses were conducted on the ERP data. First, onset latency analyses were performed. For each type of overlap (semantic or phonological), ERPs for homogeneous and heterogeneous conditions were compared by running *t*-tests on all electrodes at every sampling point (every 2 ms) from -100 to 500 ms relative to picture presentation, with the aim of identifying the latency at which the ERPs started to diverge significantly from each other. To protect against problems associated with multiple comparisons, we performed onset latency analyses using a method developed by Guthrie and Buchwald (1991; see Costa, Strijkers, Martin, & Thierry, 2009; Qu, Zhang, & Damian, 2016; Qu & Damian, 2020; Strijkers et al., 2010; Thierry, Cardebat, & Demonet, 2003 for use of this method in recent studies). This method estimates how long an interval of consecutive significant points can be expected by chance via computer simulations based on autocorrection coefficients, sample sizes, and sampling interval length. If the observed number of consecutive significant time intervals is longer than the significant interval expected by chance, this indicate a statistically significant interval, and the onset point of the consecutive significant points is taken as the onset of the corresponding effect. In this analysis, we averaged the onset latencies of those electrodes which showed significant intervals in each type, and regarded the averaged values as the onsets of the two effects.

We examined whether the onset latencies of semantics versus phonology were significantly different from each other using the jackknife approach (Miller, Patterson, & Ulrich, 1998; Ulrich & Miller, 2001). A jackknife waveform was computed for each participant i (i = 1. . . n, where n is the number of participants) by temporarily omitting participant i and computing the grand average of the difference in ERPs from the remaining n–1 participants. The jackknife onset latency was determined by the *t*-tests at every sampling point on the electrodes which showed significant intervals. To determine the statistical difference of the latencies of semantic effect and phonological effect, the onset latencies of the two effects were compared with a repeated-measures Analysis of Variance (ANOVA; type of overlap as a within-participant factor). The same procedure was conducted to compare the time points where the two effects reached their maximum value. A corrected *F*-value ($F_c = F/(n-1)^2$) was used to determine statistical significance.

Second, mean amplitude analysis was performed. Five time windows for mean amplitude analysis were selected on the basis of visible peaks in the grand average ERP waveforms, combined with the consideration of the selection of time windows in previous production studies (e.g., Costa et al., 2009; Strijkers et al., 2010): [0-180 ms], [180-250 ms] (P2), [280-320 ms] (N2), [320-380 ms] (P3), and [380-500 ms]. For each type in each time window, mean amplitudes of the selected electrodes were entered into repeated-measures ANOVA. *FDR* correction (Benjamini & Hochberg, 1995; Genovese & Wasserman, 2002) was applied on the obtained *p* values to control the potential *Type I* error.

Results

In blocked cyclic naming tasks, performance on the first cycle within an experimental block often differs from the remaining cycles. The effect of semantic blocking which is overall interfering is often facilitatory on the first cycle (e.g., Belke et al., 2005; Schnur et al., 2006; Navarrete et al., 2014), and occasionally absent (e.g., Damian & Als, 2005). This pattern has generated an extensive discussion of the possibility that semantic effects on cycle 1, and on subsequent cycles, might be caused by different underlying psychological and neural mechanisms (e.g., Belke, 2008; Navarrete, Del Prato & Mahon, 2012). Recently, Python, Fargier and Laganaro (2018) combined semantic blocking with EEG, and attributed the effect on the first cycle to postlexical interactive phonological-semantic processes and/or to self monitoring, whereas the effect on subsequent cycles reflected lexical competition. Effects of phonological overlap in blocked cyclic naming have been less well investigated, but in their pioneering study Breining et al. (2015) also suggested that effects on the first cycle might differ from the one on the remaining cycles within a block. For this reason, and in accordance with previous studies such as Breining et al., we considered only the results from cycles 2-4 for the analyses of RTs and ERPs. However, Figure 1 below shows performance results from all cycles, and analyses of behavioural and ERP results were overall similar with and without cycle 1.

Data with missing recordings (1.2%), incorrect responses (0.6%), latencies shorter than 300 ms or longer than 1500 ms (2.9%) and beyond 2.5 *SD*s (1.1%) were excluded from the behavioural and ERP analyses. For the ERP analyses, a further 8.7% of trials were excluded due to artifacts. In total, ERP analyses were based on an average of 44 segments per condition (semantic homogeneous, 43; semantic heterogeneous: 44, phonological homogeneous: 44, phonological heterogeneous: 45).

Behavioural results

As shown in Figure 1, for both types of overlap, responses in the homogeneous context were longer than that in the heterogeneous context (semantics: 562 vs. 525 ms; phonology: 531 vs. 517 ms). The linear mixed-effects model analysis showed a significant main effect of context (β = 26.18, *SE* = 5.49, t = 4.77, $p < .001$), and a marginally significant interaction between type and context (β = 22.11, *SE* = 11.57, t = 1.91, p = .056), suggesting a somewhat larger context effect in the semantic than in the phonological type of overlap. The main effect of type of overlap was not significant (semantics: 543 ms, phonology: 524 ms, β = 18.70, *SE* = 11.02, t = 1.7, p = .09). Further analyses for each type of overlap revealed a context effect for both the

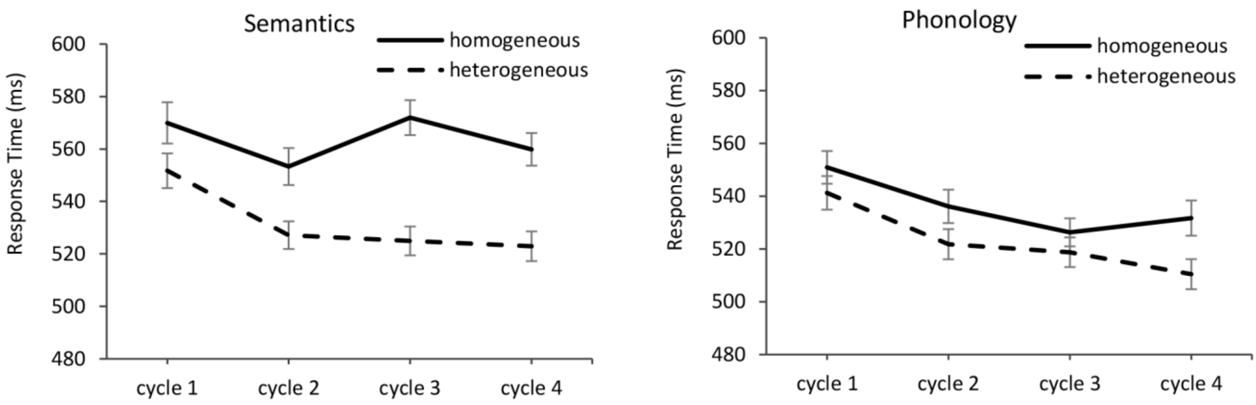

Figure 1. Mean response times by type of overlap (semantics vs. phonology) and context (homogeneous vs. heterogeneous). Error bars indicate standard error of the mean.

semantics and phonology conditions (semantics: 562 vs. 525 ms, $\beta = 37.11$, $SE = 7.56$, $t = 4.91$, $p < .001$; phonology: 531 vs. 517 ms, $\beta = 15.02$, $SE = 4.17$, $t = 3.61$, $p < .001$). Only the semantic effect interacted with cycles ($\beta = 20.57$, $SE = 10.12$, $t = 2.03$, $p = .042$) whereas the phonological effect did not ($p > .4$). No analysis was performed on accuracy as the error rate was extremely low (~0.6%).

ERP Results

Grand average ERP waveforms are displayed in Figure 2 for the semantic and phonological conditions averaged from the data of all participants on selected electrodes. Target pictures elicited an expected P1/N1/P2 ERP component in all conditions, and the homogeneous context in both conditions produced less positive amplitudes than the heterogeneous context did.

Onset latency analysis. Separately for the semantics and the phonology conditions, ERPs corresponding to the homogeneous and heterogeneous conditions were compared by running *t*-tests at every 2 ms starting -100 to 500 ms relative to picture onset over all 62 electrodes. Onset latencies were computed on averages of those electrodes in which the observed number of consecutive significant time points was larger than the critical run length in order to determine statistical significance. For the semantic effect, the averaged splitting point computed from individual splitting point estimates (9 electrodes: CP1, CP3, P1, P2, P3, P4, PO3, PO4, PZ) was 175 ms after picture onset. The averaged splitting point for the phonological effect (6 electrodes: P1, P2, PZ, POZ, OZ, PO4) was 192 ms after picture onset. Results of the jackknife method showed that the time points at which the effects started to emerge was not significantly different between the semantic and the phonological condition ($F_{corrected} < 1$). Neither did the time points differ significantly at which the semantic and phonological effects reached their respective maximum value (semantics: 243 ms, phonology: 272 ms, $F_{\text{corrected}} < 1$).

Mean amplitude analyses. Six regions of interest (ROIs) were selected to probe the scalp distribution of ERP differences: Left-Anterior (F3, FC3, FC5), Middle-Anterior (FZ, FCZ, CZ), Right-Anterior (F4, FC4, FC6), Left-Posterior (CP3, P3, P5), Middle-Posterior (CPZ, PZ, POZ), and Right-Posterior (CP4, P4, P6). For each condition, the main results of the omnibus ANOVA, conducted separately for each of the five time intervals (0-180, 180-250, 280-320, 320-380, 380-500 ms), were as follows. Moreover, critical time windows were aggregated into a larger time range, and additional analyses were conducted on the larger time range, i.e., 180-380 ms.

In the early time window from **0 ms to 180 ms** after picture onset, for the semantics condition, only the interaction between context and laterality was significant ($F_{(2,46)} = 3.82$, $p = .03$). However, post-hoc analysis in each ROI did not show any significant context effects (all *p*s > .12). For the phonology condition, neither the main effect of context nor interactions involving context were significant (*p*s > .13), and none of ROIs showed context effect (*p*s > .35). In the P2 range (**180-250 ms**), the interaction between context, anteriority and laterality was significant in the semantics condition ($F_{(2,46)} = 5.39$, $p = .007$) and marginally significant in the phonology condition ($F_{(2,46)} = 2.75$, $p = .077$). Critically, *post-hoc* analysis revealed significant semantic effects in left posterior ($t_{(23)} = -3.59$, $p = .002$), middle posterior ($t_{(23)} = -3.69$, $p = .002$), and right posterior ($t_{(23)} = -2.49$, $p = .02$), and a significant phonological effect in the middle posterior region ($t_{(23)} = -3.26$, $p = .01$). In the following N2 range (**280-320 ms**), the interaction between context, anteriority and laterality was significant in the semantics condition ($F_{(2,46)} = 3.51$, $p = .040$) and the phonology condition ($F_{(2,46)} = 3.82$, $p = .03$). The context effects were marginally significant in middle posterior ($t_{(23)} = -2.44$, $p = .069$) and right posterior ($t_{(23)} = -1.97$, $p = .091$) in the semantics condition, and middle posterior in the phonology condition ($t_{(23)} = -2.52$, $p = .058$). In the P3 range (**320-380 ms**), the interaction between context, anteriority and laterality was marginally significant in the semantics condition ($F_{(2,46)} = 2.76$, $p = .075$) and significant in the phonology condition ($F_{(2,46)} = 3.39$, $p = .044$). Left posterior region in the semantics condition ($t_{(23)} = -2.92$, $p = .023$) and middle posterior region in the phonology condition ($t_{(23)} = -2.77$, $p = .032$) demonstrated significant context effects. In the late time window (**380-500 ms**), neither the main effect of context, nor interactions involved context were significant (*p*s > 0.1), and none of ROIs showed context effect (*p*s > 0.35). Hence, both semantic and phonological relatedness showed modulation of ERPs in the same time windows, starting about 180 ms and ending at about 380 ms after picture onset.

This pattern of ERP effects was confirmed via additional analyses on a larger time window, i.e., **180-380 ms**. In the semantics condition, the interaction among anteriority, laterality and context was significant ($F_{(2,46)} = 5.56$, $p = .007$). Further analysis at each ROI (*p*-values FDR corrected) revealed a significant semantic effect in the left posterior, $t_{(23)} = -3.15$, $p = .013$, middle posterior, $t_{(23)} = -2.83$, $p = .014$, and right posterior regions, $t_{(23)} = -2.07$, $p = .049$. Similarly, the phonology condition also demonstrated a three-way interaction among anteriority, laterality and context ($F_{(2,46)} = 3.56$, $p = .036$). The phonology effect was significant in the middle posterior region ($t_{(23)} = -3.06$, $p = .016$). This pattern of ERP effects were further confirmed via additional analyses on a larger set of 18 electrodes of the central-posterior region (CZ, CPZ, PZ, POZ, C1, CP1, P1, C2, CP2, P2, C3, CP3, P3, PO3, C4, CP4, P4, PO4) [semantics: $t_{(23)} = -3.68$, $p < .001$; phonology: $t_{(23)} = -2.36$, $p = .013$]. Hence, ERP data showed that semantic overlap and phonological overlap modulated ERPs in a time window of 180-380 ms, across the central-posterior regions.

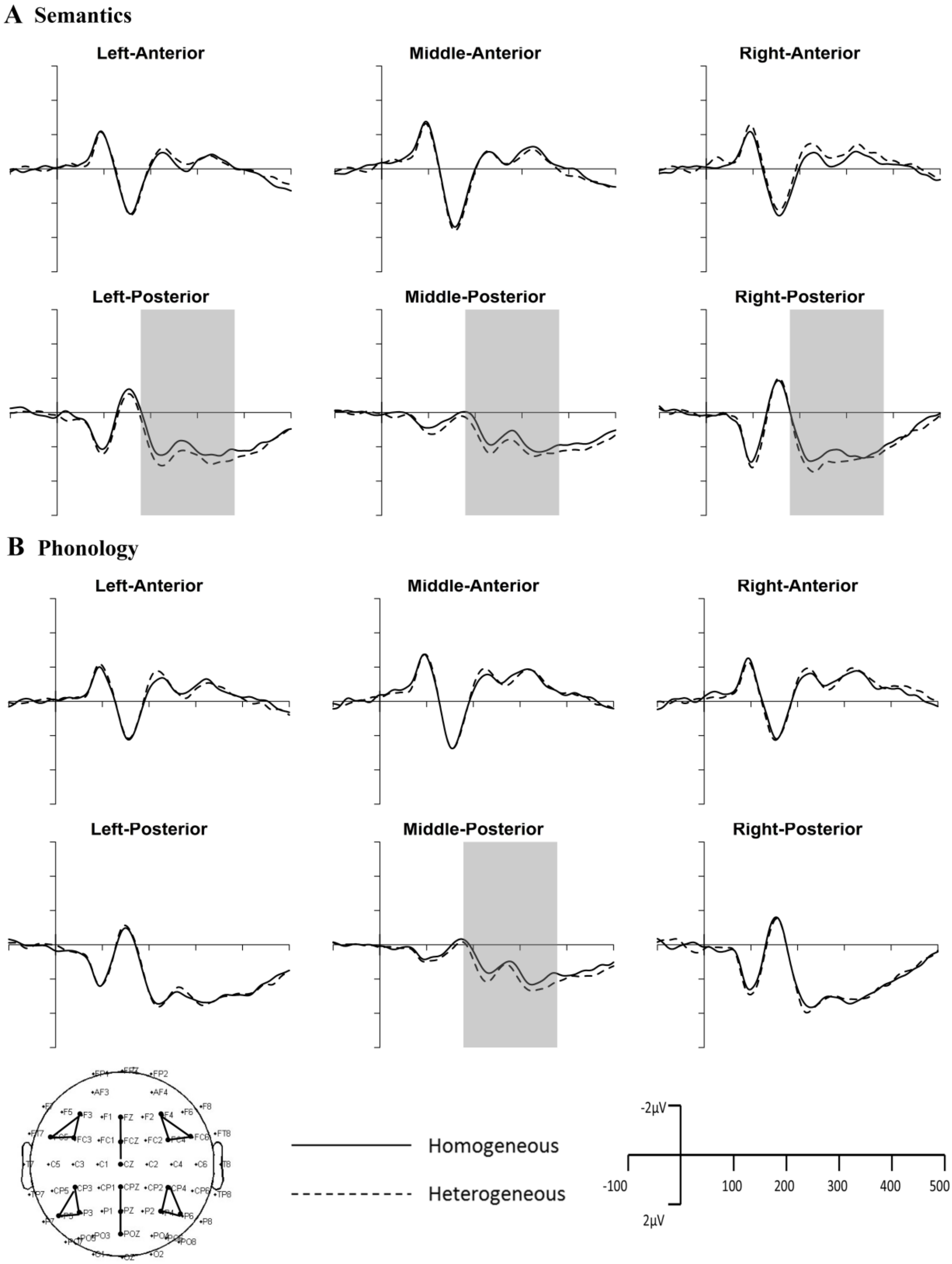

Figure 2. (A) Grand average ERPs for homogeneous (solid lines) and heterogeneous (dash lines) context in the semantic type of overlap. (B) Grand average ERPs for homogeneous (solid lines) and heterogeneous (dash lines) context in the phonological type of overlap. The onset of a target object is represented by 0 ms. Semantic and phonological overlap modulated ERPs in the time window of 180-380 ms.

Discussion

With the EEG technique, we investigated the time course of semantic and phonological processes in spoken word production via the blocked cyclic naming paradigm. Objects were embedded within lists of same-category items, or lists of phonologically overlapping items (or, within lists of unrelated items). Consistent with previous studies (Belke et al., 2005; Damian et al., 2001), objects were named more slowly in the context of same-category items than when stemming from various semantic categories. Phonological overlap in our study could occur either in first or second word position of responses within a homogeneous block, and this also resulted in slower naming responses than in the corresponding unrelated baseline condition. This latter finding – phonological interference – poses an interesting contrast with numerous previous studies in which word-initial phonological overlap was shown to result in facilitation (e.g., Meyer 1990), and it replicates the finding of Breining et al. (2016) with English materials, as well as our own (Qu, Feng, & Damian, under review) with Chinese stimuli.

Most importantly for present purposes, the ERP data showed that semantic and phonological overlap modulated ERPs in similar time windows (180-380 ms). Precise temporal analysis revealed that the semantic effect emerged with an onset of 175 ms and the phonological effect had an onset of 192 ms. The two effects did not differ significantly regarding the time points where the effects began, nor where they reached maximum. Semantic relatedness evoked less positive ERPs compared to the baseline condition, which is consistent with previous studies (Anders, Alario, Llorens, Dubarry, Trébuchon, & Liegeois-Chauvel, 2019; Janssen, Hernández-Cabrera, van der Meij, & Barber, 2015; Maess, Friederici, Damian, Meyer, & Levelt, 2002; Python, Fargier, & Laganaro, 2017; Wang, Shao, Chen, & Schiller, 2018; but see Janssen, Carreiras, & Barber, 2011). Similarly, ERPs were also more positive in the phonologically related context than in the baseline condition.

The most striking aspect of our EEG results is that onset latencies, as well as the time interval, for the two types of context effects were almost identical. Evidently, semantic processing takes place in parallel with phonological encoding, within approximately 180 ms after picture onset. This finding appears to contradict models of word production which conceive of semantic and phonological encoding as discrete processing stages (Levelt et al., 1999; the seriality assumption is also critical to Indefrey’s 2002, and Indefrey and Levelt’s 2004, time estimates of processing stages in spoken production). According to these models, lexical-semantic activation precedes phonological encoding, so in an EEG study the time window which indexes lexical-semantics should precede the one associated with access to phonology. There are several potential explanations for the empirical pattern. One possibility is that activation strongly cascades throughout the network, hence the phonological encoding system receives activation almost as soon as processing begins at the lexical-semantic stage. Therefore, our finding that semantic and phonological effects emerged at the similar time interval with slightly earlier semantic effect might be compatible with models of lexical access which incorporate cascading activation (and potentially, feedback) between lexical-semantic and phonological encoding (Dell, 1986; Dell & O’Seaghda, 1992; Peterson & Savoy, 1998; Rapp & Goldrick, 2000).

An alternative and new perspective is offered by neural assembly models where all components of speech production are organised in a single functional unit in the way of connected neural populations (Pulvermüller, 1999; Strijkers, 2016; Strijkers & Costa, 2016). These components could be activated synchronously, or near-simultaneously. A critical difference between the cascading and neural assembly models concerns whether the interaction is restricted to adjacent levels. The cascading models predict that the cascading activation only emerges between adjacent stages (e.g., from visual processing to semantics; from semantics to phonology) whereas the non-hierarchical neural assembly models predict that interactivity likely reaches beyond adjacent layers (e.g., from visual processing to phonology). In the present study, the semantic effect and phonological effects are associated with two neighboring processing levels (lexical-semantic and phonological encoding) and thus our results are not able to distinguish between the two types of models. However, it should be noted that cascaded interactive processing models generally assume that adjacent processing stages are separated by functional delays of roughly 100 ms (Dell & O'Seaghdha, 1992). The tiny ~20 ms of onset latency difference between semantics and phonology is not quantitatively compatible with this assumption. Therefore, the current findings are not easily explained by cascading models and instead favor neural assembly models.

The time window that hosts semantic effects in our study (180-380 ms) is compatible with findings from ERP studies with the blocked cyclic naming paradigm (Aristei, Melinger, & Abdel Rahman, 2011; Maess et al., 2002; Python et al., 2017). In addition, this time course is also consistent with studies using other overt picture naming tasks. In a picture naming study, Strijkers et al. (2010) observed a “word frequency effect” (an effect which is assumed to arise from lexical-semantic retrieval) which started at 180 ms after picture presentation. Miozzo, Pulvermüller and Hauk (2014) used a multiple linear regression approach to MEG analysis and found early effects (around 150 ms) of variables related to semantic processing. It has been shown that the time course of semantic processing is independent of participants’ response speed (Laganaro Valente, & Perret, 2012) and naming modalities, i. e., speaking vs. writing (Perret & Laganaro, 2012). The onset of the semantic effect found in the present study (175 ms) is broadly in agreement with the estimated onset time of lexical-semantic encoding proposed by Indefrey and Levelt (2004; 175 ms) and the update provided by Indefrey (2011; 200 ms).

By contrast, the onset of the phonological effect in our study (192 ms) is earlier than predicted from the estimates by Indefrey and Levelt (250 ms) and Indefrey (275 ms). However, such an early onset is broadly in line with recent findings from related tasks which have suggested that phonology is accessed around 200 ms (Miozzo, et al., 2015; Qu et al., 2012; Qu & Damian, 2020; Strijkers et al., 2017; Wang, Wong, Wang, & Chen, 2017; Yu et al., 2014; Zhang, Wong, Wang, & Chen, 2018). For example, overlap of a single phoneme between successive words modulated ERPs in a time window of 200-300 ms in a coloured picture naming task (Qu et al., 2012) and 180-300 ms in a picture-picture priming task (Yu et al. 2014). However, it is worth noting that EEG results from the picture-word interference task have suggested a later time course of phonological encoding. For instance, Zhu, Damian, and Zhang (2015) and Wong, Wang, Ng, and Chen (2016) found phonologically based ERP effects at 450-600 ms and 500-600 ms respectively; Dell’Acqua et al. (2010) observed form-based ERP effects in a slightly earlier time interval of 250-400 ms after picture onset, but this interval is still later than that found in the present study. One possible reason for this discrepancy might be that in picture-word experiments, processing of a distractor word itself delays the normally rapid phonological encoding of the spoken utterance.

How does the semantic context effect relate to spoken word production? Interfering effects of semantic context in the blocked cyclic naming paradigm were initially interpreted via competitive selection among a set of words which was co-activated by the spreading activation mechanism (Damian et al., 2001). However, a subsequent study (Damian & Als, 2005) showed that the effect did not diminish when several unrelated items intervened between critical naming responses. A semantic context effect which persists over a relatively long time period is incompatible with an explanation of the effect based on spreading activation, which is generally assumed to be short-lived. An alternative account (Damian & Als, 2005; Howard et al., 2006; Oppenheim et al., 2010) is that engagement of the semantic-lexical pathway leads to “incremental learning", i.e. a slight but persistent increment in the connection weights between the representational layers. This will increase efficiency when the same item is named again ("repetition priming"; e.g., Cave, 1997) but at the same time it will slightly slow down the naming of semantically related items. Howard et al. (2006) presented a computational model in which semantic features were connected to nodes within a lexical layer, and the latter inhibited each other. Slight increments in the connections simulated the cumulative semantic inhibition effect. Oppenheim et al. (2010) provided an extensive computational simulation of these phenomena. They introduced a model in which semantic features were connected to lexical nodes, and semantic-to-lexical links were modified on each trial such that weights from semantically active features to target lexical nodes were increased, whereas weights to all other lexical nodes were decreased. This principle of "competitive learning" was able to generate both cumulative semantic inhibition and cyclic semantic blocking, but intriguingly, lexical competition was not necessary to do so (however, see Roelofs, 2018, for a more recent computational simulation of various semantic effects in spoken word production which retains the notion of competitive lexical selection as a central architectural feature).

The present study found that just like semantic overlap, phonological form overlap across different word positions caused interference, hence replicating a pattern first demonstrated by Breining et al. (2016). In a recent study of ours, we (Qu, Feng, & Damian, under review) have demonstrated that phonological interference was found even when (semantically and phonologically) unrelated filler pictures were interleaved with the critical targets. Hence, the phonological interference effect is reasonably long-lasting. For this reason, we interpret this finding as implying that the phonological effect is also due to competitive incremental learning (as was claimed for the semantic effect; see above), but incremental learning occurs between lexical and phonological representations for phonology. According to this mechanism, object naming elicits a slight and persistent modification of connections between lexical and phonological representations. For example, consider a mini-scenario with only two phonologically related objects ("cat"-"cap"). On trial N-1, the lexical code for "cat" and its corresponding phonological codes are activated. Via feedback from the phonological to the lexical level, phonologically related lexical codes (e.g., "cap") will be co-activated. Following the trial N-1, all links between the target code and the phonological level are strengthened, and connections between the lexical node "cap" and the phonological level are decremented. On a subsequent trial, if the phonological related item "cap" is to be named, some of the connections between the lexical node and the phonological codes have been already weakened on the previous trial, and so naming times and/or accuracy for a naming response to "cap" will be detrimentally affected. Speculatively, incremental learning constitutes a universal principle in semantic-to-lexical and lexical-to-phonology mappings.

As far as ERP components are concerned, as reported previously for word frequency, cognate status (Strijkers, Costa, & Thierry, 2010; see also Strijkers et al., 2013) and cumulative semantic interference in picture naming (Costa et al., 2009), we found that both the semantic effect and the phonological effect elicited the same modulations of electrophysiological responses at P2, N2 and P3. Perhaps the most interesting ERP component is P2, which appears to be modulated by various lexical variables such as lexical frequency, cognate status and semantic interference and is thus assumed to reflect lexical access, with larger amplitudes associated with less accessible representations (e.g., low frequency words and non-cognates). Similarly, our study reported the same directionality of effects: larger P2 amplitudes for homogeneous conditions (slower RTs) as compared with heterogeneous conditions (faster RTs). Based on these findings, we emphasize the effectiveness of the P2 component as an index of lexical access in speech production.

In sum, our behavioural results showed that both semantic and phonological overlap between spoken response words within an experimental block of picture naming responses produced longer naming latencies, compared to the baseline conditions. Our ERP results provide evidence for the claim that lexical-semantic and phonological processing proceed largely in parallel. These findings will provide important constraints on models of spoken word production.

Funding

This work was supported by the National Natural Science Foundation of China (No. 31771212 and No. 61621136008), Youth Innovation Promotion Association (Chinese Academy of Sciences), and Youth Talent Project (China Association for Science and Technology) to Qu.

Notes

The authors declare that there are no conflicts of interest.

References

Abdel Rahman, R. & Melinger, A. (2007). When bees hamper the production of honey: Lexical interference from associates in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 604-614. DOI: https://doi.org/10.1037/0278-7393.33.3.604

Abdel Rahman, R. & Melinger, A. (2009). Semantic context effects in language production: A swinging lexical network proposal and a review. *Language and Cognitive Processes*, *24*, 713-734. DOI: https://doi.org/10.1080/01690960802597250

Abdel Rahman, R. & Sommer, W. (2003). Does phonological encoding in speech production always follow the retrieval of semantic knowledge? Electrophysiological evidence for parallel processing. *Cognitive Brain Research*, *16*, 372-382. DOI: https://doi.org/10.1016/S0926-6410(02)00305-1

Anders, R., Alario, F. X., Llorens, A., Dubarry, A. S., Trébuchon, A. & Liegeois-Chauvel, C. (2019). Cortical dynamics of semantic priming and interference during word production: An intracerebral study. *Journal of Cognitive Neuroscience*, *31*, 978-1001. DOI: https://doi.org/10.1162/jocn_a_01406

Aristei, S., Melinger, A. & Abdel Rahman, R. (2011). Electrophysiological chronometry of semantic context effects in language production. *Journal of Cognitive Neuroscience*, *23*, 1567-1586. DOI: https://doi.org/10.1162/jocn.2010.21474

Baayen, H., Bates, D., Kliegl, R. & Vasishth, S. (2015). RePsychLing: Data sets from psychology and linguistics experiments. R package version 0.0, 4.

Baayen, R. H., Davidson, D. J. & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390-412. DOI: https://doi.org/10.1016/j.jml.2007.12.005

Bates, D., Kliegl, R., Vasishth, S. & Baayen, H. (2015). Parsimonious mixed models. *arXiv preprint* arXiv:1506.04967.

Bates, D. & Maechler, M. (2009). Ime4: Linear mixed-effects models using S4 classes. R package version 0.999375-31. CRAN. R-project. ore/packaee= lme4.

Belke, E., Meyer, A. S. & Damian, M. F. (2005). Refractory effects in picture naming as assessed in a semantic blocking paradigm. *The Quarterly Journal of Experimental Psychology Section A, 58, 667-692.* DOI: https://doi.org/10.1080/02724980443000142

Benjamini, Y. & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological), 57*, 289-300. DOI: https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Breining, B., Nozari, N. & Rapp, B. (2016). Does segmental overlap help or hurt? Evidence from blocked cyclic naming in spoken and written production. *Psychonomic Bulletin & Review, 23,* 500-506. DOI: https://doi.org/10.3758/s13423-015-0900-x

Camen, C., Morand, S. & Laganaro, M. (2010). Re-evaluating the time course of gender and phonological encoding during silent monitoring tasks estimated by ERP: serial or parallel processing? *Journal of Psycholinguistic Research*, *39*, 35-49. DOI: https://doi.org/10.1007/s10936-009-9124-4

Cave, C. B. (1997). Very long-lasting priming in picture naming. *Psychological Science*, *8*, 322-325. DOI: https://doi.org/10.1111/j.1467-9280.1997.tb00446.x

Chen, T. M. & Chen, J. Y. (2013). The syllable as the proximate unit in Mandarin Chinese word production: An intrinsic or accidental property of the production system? *Psychonomic Bulletin & Review*, *20*, 154-162. DOI: https://doi.org/10.3758/s13423-012-0326-7

Costa, A., Strijkers, K., Martin, C. & Thierry, G. (2009). The time course of word retrieval revealed by event-related brain potentials during overt speech. *Proceedings of the National Academy of Sciences*, *106*, 21442-21446. DOI: https://doi.org/10.1073/pnas.0908921106

Damian, M. F. & Als, L. C. (2005). Long-lasting semantic context effects in the spoken production of object names. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 1372-1384. DOI: https://doi.org/10.1037/0278-7393.31.6.1372

Damian, M. F. & Bowers, J. S. (2003). Effects of orthography on speech production in a form-preparation paradigm. *Journal of Memory and Language*, *49*, 119-132. DOI: https://doi.org/10.1.1.5.2255

Damian, M. F. & Martin, R. C. (1999). Semantic and phonological codes interact in single word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 345-361. DOI: https://doi.org/10.1037//0278-7393.25.2.345

Damian, M. F., Vigliocco, G. & Levelt, W. J. (2001). Effects of semantic context in the naming of pictures and words. *Cognition*, *81*, B77-B86. DOI: https://doi.org/10.1016/S0010-0277(01)00135-4

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*, 283. DOI: https://doi.org/10.1037/0033-295X.93.3.283

Dell, G. S. & O'Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition*, *42*, 287-314. DOI: https://doi.org/10.1016/0010-0277(92)90046-K

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M. & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review*, *104*, 801-838. DOI: https://doi.org/10.1037/0033-295X.104.4.801

Dell'Acqua, R., Sessa, P., Peressotti, F., Mulatti, C., Navarrete, E. & Grainger, J. (2010). ERP evidence for ultra-fast semantic processing in the picture–word interference paradigm. *Frontiers in Psychology*, 1: 177. DOI: https://doi.org/10.3389/fpsyg.2010.00177

Garrett, M. F. (1975). The analysis of sentence production. In G. Bower (Ed.), *Psychology of learning and motivation*. New York: Academic Press. Vol. 9, p.133-177. DOI: https://doi.org/10.1016/S0079-7421(08)60270-4

Genovese, C. & Wasserman, L. (2002). Operating characteristics and extensions of the false discovery rate procedure. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*, 499-517. DOI: https://doi.org/10.1111/1467-9868.00347

Glaser, M. O. & Glaser, W. R. (1982). Time course analysis of the Stroop phenomenon. *Journal of Experimental Psychology: Human Perception and Performance, 8*, 875-894. DOI: https://doi.org/10.1037/0096-1523.8.6.875

Glaser, W. R. & Düngelhoff, F. J. (1984). The time course of picture-word interference. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 640-654. DOI: https://doi.org/10.1037/0096-1523.10.5.640

Guthrie, D. & Buchwald, J. S. (1991). Significance testing of difference potentials. *Psychophysiology, 28,* 240-244. DOI: https://doi.org/10.1111/j.1469-8986.1991.tb00417.x

Howard, D., Nickels, L., Coltheart, M. & Cole-Virtue, J. (2006). Cumulative semantic inhibition in picture naming: Experimental and computational studies. *Cognition*, *100*, 464-482. DOI: https://doi.org/10.1016/j.cognition.2005.02.006

Indefrey, P. (2002). Listen und Regeln: Erwerb und Repräsentation der schwachen Substantivdeklination des Deutschen (Doctoral dissertation, Heinrich Heine Universität Düsseldorf).

Indefrey, P. (2011). The spatial and temporal signatures of word production components: a critical update. *Frontiers in Psychology*, 2: 255. DOI: https://doi.org/10.33892/fpsyg.2011.00255

Indefrey, P. & Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92,* 101-144. DOI: https://doi.org/10.1016/j.cognition.2002.06.001

Janssen, N., Carreiras, M. & Barber, H. A. (2011). Electrophysiological effects of semantic context in picture and word naming. *Neuroimage*, *57*, 1243-1250. DOI: https://doi.org/10.1016/j.neuroimage.2011.05.015

Janssen, N., Hernández-Cabrera, J. A., van der Meij, M. & Barber, H. A. (2015). Tracking the time course of competition during word production: Evidence for a post-retrieval mechanism of conflict resolution. *Cerebral Cortex*, *25*, 2960-2969. DOI: https://doi.org/10.1093/cercor/bhu092

Jung, T. P., Makeig, S., Humphries, C., Lee, T. W., Mckeown, M. J., Iragui, V. & Sejnowski, T. J. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, *37*, 163-178. DOI: https://doi.org/10.1111/1469-8986.3720163

Kroll, J. F. & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, *33*, 149-174. DOI: https://doi.org/10.1006/jmla.1994.1008

Laganaro, M. & Perret, C. (2011). Comparing electrophysiological correlates of word production in immediate and delayed naming through the analysis of word age of acquisition effects. *Brain Topography*, *24*, 19-29. DOI: https://doi.org/10.1007/s10548-010-0162-x

Laganaro, M., Valente, A. & Perret, C. (2012). Time course of word production in fast and slow speakers: a high density ERP topographic study. *NeuroImage*, *59*, 3881-3888. DOI: https://doi.org/10.1016/j.neuroimage.2011.10.082

Levelt, W. J., Roelofs, A. & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22,* 1-38. DOI: https://doi.org/10.1017/S0140525X99001776

Liu, Y., Hao, M., Li, P. & Shu, H. (2011). Timed picture naming norms for Mandarin Chinese. *PLoS One, 6,* e16505. DOI: https://doi.org/10.1371/journal.pone.0016505

Maess, B., Friederici, A. D., Damian, M., Meyer, A. S. & Levelt, W. J. (2002). Semantic category interference in overt picture naming: Sharpening current density localization by PCA. *Journal of Cognitive Neuroscience*, *14*, 455-462. DOI: https://doi.org/10.1162/089892902317361967

Mahon, B. Z., Costa, A., Peterson, R., Vargas, K. A. & Caramazza, A. (2007). Lexical selection is not by competition: a reinterpretation of semantic interference and facilitation effects in the picture-word interference paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 503-535. DOI: https://doi.org/10.1037/0278-7393.33.3.503

Meyer, A. S. (1990). The time course of phonological encoding in language production: The encoding of successive syllables of a word. *Journal of Memory and Language*, *29*, 524-545. DOI: https://doi.org/10.1016/0749-596X(90)90050-A

Meyer, A. S. (1991). The time course of phonological encoding in language production: Phonological encoding inside a syllable. *Journal of Memory and Language*, *30*, 69-89. DOI: https://doi.org/10.1016/0749-596X(91)90011-8

Miller, J., Patterson, T. & Ulrich, R. (1998). Jackknife-based method for measuring LRP onset latency differences. *Psychophysiology*, *35*, 99-115. DOI: https://doi.org/10.1111/1469-8986.3510099

Miozzo, M., Pulvermüller, F. & Hauk, O. (2015). Early parallel activation of semantics and phonology in picture naming: Evidence from a multiple linear regression MEG study. *Cerebral Cortex*, *25*, 3343-3355. DOI: https://doi.org/10.1093/cercor/bhu137

Mirman, D. (2014). Growth curve analysis and visualization using R. Boca Raton, FL: Chapman & Hall/ CRC Press. DOI: https://doi.org/10.1201/9781315373218

Munding, D., Dubarry, A. S. & Alario, F. X. (2016). On the cortical dynamics of word production: A review of the MEG evidence. *Language, Cognition and Neuroscience*, *31*, 441-462. DOI: https://doi.org/10.1080/23273798.2015.1071857

O'Séaghdha, P. G. & Frazer, A. K. (2014). The exception does not rule: Attention constrains form preparation in word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 797-810. DOI: https://doi.org/10.1037/a0035576

O'Seaghdha, P. G., Chen, J.-Y. & Chen, T.-M. (2010). Proximate units in word production: Phonological encoding begins with syllables in Mandarin Chinese but with segments in English. *Cognition, 115,* 282-302. DOI: https://doi.org/10.1016/j.cognition.2010.01.001

Oppenheim, G. M., Dell, G. S. & Schwartz, M. F. (2010). The dark side of incremental learning: A model of cumulative semantic interference during lexical access in speech production. *Cognition, 114*, 227-252. DOI: https://doi.org/10.1016/j.cognition.2009.09.007

Perret, C. & Laganaro, M. (2012). Comparison of electrophysiological correlates of writing and speaking: a topographic ERP analysis. *Brain Topography, 25*, 64-72. DOI: https://doi.org/10.1007/s10548-011-0200-3

Peterson, R. R. & Savoy, P. (1998). Lexical selection and phonological encoding during language production: Evidence for cascaded processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 539-557. DOI: https://psycnet.apa.org/doi/10.1037/0278-7393.24.3.539

Pulvermüller, F. (1999). Words in the brain's language. *Behavioral and Brain Sciences*, 22(2), 253-279. DOI: https://doi.org/10.1017/S0140525X9900182X

Python, G., Fargier, R. & Laganaro, M. (2018). ERP evidence of distinct processes underlying semantic facilitation and interference in word production. *Cortex*, 99, 1-12. DOI: https://doi.org/10.1016/j.cortex.2017.09.008

Qu, Q. & Damian, M. F. (2020). An electrophysiological analysis of the time course of phonological and orthographic encoding in written word production. *Language, Cognition and Neuroscience*, 35(3), 360-373. DOI: https://doi.org/10.1080/23273798.2019.1659988

Qu, Q., Damian, M. F. & Kazanina, N. (2012). Sound-sized segments are significant for Mandarin speakers. *Proceedings of the National Academy of Sciences*, *109*(35), 14265-14270. DOI: https://doi.org/10.1073/pnas.1200632109

Qu, Q., Feng, C. & Damian, M.F. Under review. Interference effects of phonological similarity in word production arise from competitive incremental learning.

Qu, Q., Feng, C., Hou, F. & Damian, M. F. (2020). Syllables and phonemes as planning units in Mandarin Chinese spoken word production: Evidence from ERPs. *Neuropsychologia*, 107559. DOI: https://doi.org/10.1016/j.neuropsychologia.2020.107559

Qu, Q., Zhang, Q. & Damian, M. F. (2016). Tracking the time course of lexical access in orthographic production: An event-related potential study of word frequency effects in written picture naming. *Brain and Language*, *159*, 118-126. DOI: https://doi.org/10.1016/j.bandl.2016.06.008

Rapp, B. & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review*, *107*(3), 460. DOI: https://psycnet.apa.org/doi/10.1037/0033-295X.107.3.460

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, *42*, 107-142. DOI: https://doi.org/10.1016/0010-0277(92)90041-F

Roelofs, A. (2018). A unified computational account of cumulative semantic, semantic blocking, and semantic distractor effects in picture naming. *Cognition, 172*, 59-72. DOI: https://doi.org/10.1016/j.cognition.2017.12.007

Schmitt, B. M., Münte, T. F. & Kutas, M. (2000). Electrophysiological estimates of the time course of semantic and phonological encoding during implicit picture naming. *Psychophysiology*, *37*, 473-484. DOI: https://doi.org/10.1111/1469-8986.3740473

Schriefers, H., Meyer, A. S. & Levelt, W. J. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language*, *29*, 86-102. DOI: https://doi.org/10.1016/0749-596X(90)90011-N

Starreveld, P. A. & La Heij, W. (1995). Semantic interference, orthographic facilitation, and their interaction in naming tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 686-698. DOI: https://doi.org/10.1037/0278-7393.21.3.686

Starreveld, P. A. & La Heij, W. (1996). Time-course analysis of semantic and orthographic context effects in picture naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 869-918. DOI: https://doi.org/10.1037/0278-7393.22.4.896

Strijkers, K. (2016). A neural assembly–based view on word production: The bilingual test case. *Language Learning*, *66*(S2), 92-131. DOI: https://doi.org/10.1111/lang.12191

Strijkers, K. (2016). Can hierarchical models display parallel cortical dynamics? A non-hierarchical alternative of brain language theory. *Language, Cognition and Neuroscience*, *31*, 465-469. DOI: https://doi.org/10.1080/23273798.2015.1096403

Strijkers, K., Baus, C., Runnqvist, E., FitzPatrick, I. & Costa, A. (2013). The temporal dynamics of first versus second language production. *Brain and Language, 127,* 6–11. DOI: https://doi.org/10.1016/j.bandl.2013.07.008

Strijkers, K., Costa, A. (2011). Riding the lexical speedway: A critical review on the time course of lexical selection in speech production. *Frontiers in Psychology*, *2*, 356. DOI: https://doi.org/10.3389/fpsyg.2011.00356

Strijkers, K., Costa, A. (2016). The cortical dynamics of speaking: Present shortcomings and future avenues. *Language, Cognition and Neuroscience*, *31*, 484-503. DOI: https://doi.org/10.1080/23273798.2015.1120878

Strijkers, K., Costa, A. & Pulvermüller, F. (2017). The cortical dynamics of speaking: Lexical and phonological knowledge simultaneously recruit the frontal and temporal cortex within 200 ms. *NeuroImage*, *163,* 206-219. DOI: https://doi.org/10.1016/j.neuroimage.2017.09.041

Strijkers, K., Costa, A. & Thierry, G. (2010). Tracking lexical access in speech production: electrophysiological correlates of word frequency and cognate effects. *Cerebral Cortex*, *20*, 912-928. DOI: https://doi.org/10.1093/cercor/bhp153

Taylor, J. K. & Burke, D. M. (2002). Asymmetric aging effects on semantic and phonological processes: naming in the picture-word interference task. *Psychology and Aging*, *17*, 662-676. DOI: https://doi.org/10.1037//0882-7974.17.4.662

Thierry, G., Cardebat, D. & Demonet, J. F. (2003). Electrophysiological comparison of grammatical processing and semantic processing of single spoken nouns. *Cognitive Brain Research*, *17*, 535-547. DOI: https://doi.org/10.1016/S0926-6410(03)00168-X

Ulrich, R. & Miller, J. (2001). Using the jackknife-based scoring method for measuring LRP onset effects in factorial designs. *Psychophysiology*, *38*, 816-827. DOI: https://doi.org/10.1111/1469-8986.3850816

Van Turennout, M., Hagoort, P. & Brown, C. M. (1997). Electrophysiological evidence on the time course of semantic and phonological processes in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 787-806. DOI: https://doi.org/10.1037//0278-7393.23.4.787

Wang, J., Wong, A. W. K., Wang, S. & Chen, H. C. (2017). Primary phonological planning units in spoken word production are language-specific: Evidence from an ERP study. *Scientific Reports, 7*, 1-8. DOI: https://doi.org/10.1038/s41598-017-06186-z

Wang, M., Shao, Z., Chen, Y. & Schiller, N. O. (2018). Neural correlates of spoken word production in semantic and phonological blocked cyclic naming. *Language, Cognition and Neuroscience*, *33*, 575-586. DOI: https://doi.org/10.1080/23273798.2017.1395467

Wong, A. W. K., Wang, J., Ng, T. Y. & Chen, H. C. (2016). Syllabic encoding during overt speech production in Cantonese: Evidence from temporal brain responses. *Brain Research*, *1648*, 101-109. DOI: https://doi.org/10.1016/j.brainres.2016.07.032

Yu, M., Mo, C. & Mo, L. (2014). The role of phoneme in Mandarin Chinese production: Evidence from ERPs. *PloS One*, 9(9): e106486. DOI: https://doi.org/10.1371/journal.pone.0106486

Zhang, Q., Yu, B., Zhang, J., Jin, Z. & Li, L. (2018). Probing the timing recruitment of Broca's area in speech production for Mandarin Chinese: A TMS study. *Frontiers in Human Neuroscienc*e, *12*, 1–8. DOI: https://doi.org/10.3389/fnhum.2018.00133

Zhu, X., Damian, M. F. & Zhang, Q. (2015). Seriality of semantic and phonological processes during overt speech in Mandarin as revealed by event-related brain potentials. *Brain and language*, 144, 16-25. DOI: https://doi.org/10.1016/j.bandl.2015.03.007

Appendix. Materials used in the experiment

		Heterogeneous Lists			
Semantics	Homogeneous Lists	摩托 (mo2tuo2) (motorbike)	飞机 (fei1ji1) (airplane)	轮船 (lun2chuan2) (steamboat)	轿车 (jiao4che1) (car)
		鼻子 (bi2zi0) (nose)	眼睛 (yan3jing1) (eye)	手指 (shou3zhi3) (finger)	耳朵 (er3duo0) (ear)
		樱桃 (ying1tao2) (cherry)	菠萝 (bo1luo2) (pineapple)	西瓜 (xi1gua1) (watermelon)	柿子 (shi4zi0) (persimmon)
		围巾 (wei2jin1) (muffler)	裤子 (ku4zi0) (pants)	衬衫 (chen4shan1) (shirt)	背心 (bei4xin1) (vest)
Phonology	Homogeneous Lists	圣经 (sheng4**jing**1) (bible)	镜子 (**jing**4zi0) (mirror)	鲸鱼 (**jing**1yu2) (whale)	水井 (shui3**jing**3) (water well)
		电视 (dian4**shi**4) (television)	试管 (**shi**4guan3) (test tube)	石头 (**shi**2tou2) (stone)	钥匙 (yao4**shi**0) (key)
		屋檐 (wu1**yan**2) (eave)	燕子 (**yan**4zi0) (swallow)	颜料 (**yan**2liao4) (paint)	香烟 (xiang1**yan**1) (cigarette)
		熊猫 (xiong2**mao**1) (panda)	毛巾 (**mao**2jin1) (towel)	帽子 (**mao**4zi0) (cap)	长矛 (chang2**mao**2) (spear)

Note: The number denotes the tone for the preceding syllable. There are four tones in Mandarin. The number 0 represents a neutral tone. The overlapped syllable is underlined.